



King's Research Portal

DOI:

[10.1049/iet-com.2018.5300](https://doi.org/10.1049/iet-com.2018.5300)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Al-Salih, H., Nakhai, M. R., & Le, T. A. (2018). Enhanced sparse Bayesian learning-based channel estimation with optimal pilot design for massive MIMO-OFDM systems. *IET Communications*, 12(17), 2174-2180. <https://doi.org/10.1049/iet-com.2018.5300>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Enhanced Sparse Bayesian Learning-based Channel Estimation for Massive MIMO-OFDM Systems

Hayder AL-Salihi*, Mohammad Reza Nakhai* and Tuan Anh Le⁺

*Centre for Telecommunications Research, Department of Informatics, King's College London, London, UK

⁺Department of Design Engineering and Mathematics, Middlesex University, London, UK

Email: ana.al-salihi@kcl.ac.uk

Abstract—Pilot contamination limits the potential benefits of massive multiple input multiple output (MIMO) systems. To mitigate pilot contamination, in this paper, an efficient channel estimation approach is proposed for massive MIMO systems, using sparse Bayesian learning (SBL) namely coupled hierarchical Gaussian framework where the sparsity of each coefficient is controlled by its own hyperparameter and the hyperparameters of its immediate neighbours. The simulation results show that the proposed method can reconstruct original channel coefficients more effectively compared to the conventional channel estimators in terms of channel estimation accuracy in the presence of pilot contamination.

Index Terms—Sparse Bayesian learning; massive MIMO; channel estimation.

I. INTRODUCTION

THE main activity of recent research has identified that the major targets for the next generation of mobile communications, the so-called fifth generation of mobile communications, are to achieve 1000 times the system capacity and 10 times the spectral efficiency, energy efficiency and data rate, and 25 times the average cell throughput [1]. From a high-level perspective, there is a promising technology that enables reaching higher fifth generation targets, called a massive multiple input multiple output (MIMO). A massive MIMO can be defined as a system using a large number of antennas at the base station; accordingly, a significant beamforming can be achieved and the system capacity can serve a large number of users [2].

When comparing massive MIMO to the conventional MIMO systems, massive MIMO show several advantageous aspects. Firstly, as the number of the antennas at the base station goes to high values, the simplest coherent combiner and linear precoder turn out to be optimal. Secondly, by exploiting the features of the channel reciprocity, additional antennas increase the network capacity significantly without the need for additional feedback overhead. Thirdly, enabling the power reduction in the uplink and in the downlink can provide the potential for small-cell size shrinking [3].

mmWave??

Meanwhile, orthogonal frequency division multiplexing (OFDM) can also improve system performance and increase

spectral efficiency under frequency selective channels. Therefore, integration of both massive MIMO and OFDM technologies, which is usually denoted as massive MIMO-OFDM, is essential to enhance system capacity. Massive MIMO-OFDM is becoming a promising alternative among the upcoming 5G technologies.

The major limiting factor in massive MIMO is the availability of accurate, instantaneous channel state information (CSI) at the base station. The CSI is typically acquired by transmitting predefined pilot signals and estimating the channel coefficients from the received signals by applying an appropriate estimation algorithm [1]-[3].

Channel estimation accuracy depends on having perfect orthogonal pilots allocated to the users; however, to achieve high spectral efficiency, the same carrier frequency should be used in the neighbouring cells by following a specific reuse pattern. This leads to the creation of a spatially correlated inter-cell interference, known as pilot contamination, which reduces the estimation performance and spectral efficiency [1]-[3].

The pilot contamination problem was analyzed in [4] and it has shown that the precoding downlink signal of the base station in the serving cell contaminated the received signal of the users roaming in other cells. The authors of [5] analyzed the pilot contamination problems in multi-cell massive MIMO systems relying on a large antennas at the base station, and demonstrated that the pilot contamination problem persisted in large-scale MIMO [6].

However, pilot contamination could be reduced by reducing the number of pilots. A multi-user scenario therefore needs to reduce the number of pilots without affecting the channel impulse response (CIR) quality. Hence, the development of efficient channel estimation techniques for massive MIMO that are computationally less complex and require a fewer number of pilots is a challenge that should be thoroughly addressed [7].

Compressed sensing (CS) techniques have received much attention since they can recover unknown signals from just a small number of measurements, thereby using significantly fewer samples than is possible via the conventional Nyquist rate, which is the signal recovery scheme developed for CS to exploit the sparse nature of signals (that is, only a small number of components in a signal vector are non-zero). CS allows for accurate system parameter estimation with less

*This work is supported by Iraqi Higher Committee of Educational Development (HCED).

training, thereby addressing the pilot contamination problem and improving bandwidth efficiency as a consequence [4], [5]. However, classical CS algorithms require prior knowledge of channel sparsity, which is usually unknown in practical scenarios [4]-[6]. In addition, to applying CS algorithms, the sampling matrix must satisfy the restricted isometry property (RIP) to guarantee reliable estimation. Such a condition cannot be easily verified because it is computationally demanding [6], [7].

Recently, several methods has been proposed to tame the scarcity of CS-based channel estimation, however, these works assume dependency between antenna elements, as in realistic environments MIMO channel is generally correlated and statistically dependent, as the antennas are not sufficiently separated and the propagation environment does not provide a sufficient amount of rich scattering [8].

Considering the impact of antenna correlation, in this paper, we proposed an improved channel estimation technique based on sparse Bayesian learning (SBL) scheme, namely, a pattern-coupled hierarchical Gaussian framework [9]-[10], whereby, a priori probabilistic information regarding channel sparsity and the feature of the sparsity coefficients are being controlled by its own hyperparameter, and its neighbouring hyperparameters can be exploited for more reliable channel recovery to mitigate the pilot contamination problem. Also, the sampling matrix condition is efficiently overcome based on probabilistic formulation. We have also proposed enhancing the performance of the SBL-based estimator through the principle of thresholding to select the most significant taps to improve channel estimation accuracy. Furthermore, the cramer Rao bound (CRB) has also derived as a reference line.

The remainder of this paper is organized as follows. The multi-cell massive MIMO system model is presented in Section II. The SBL-based Channel Estimator is analyzed in Section III. Section IV presents the simulation results and the conclusions are drawn in Section V.

The following notations are adopted throughout this paper: for any matrix \mathbf{A} , $A_{i,j}$ denotes the (i,j) th element, while the superscripts $(.)^T$, $(.)^{-1}$ and $(.)^H$ denote the transpose operator, the inverse operator and the conjugate transpose operator, respectively. $\text{tr}(\cdot)$ denotes the trace operator. A diagonal matrix with a_1, \dots, a_N on the main diagonal is denoted $\text{diag}(a_1, \dots, a_N)$. Bold font is used to denote matrices and vectors, lower and upper case represents the time domain and frequency domain, respectively. The Frobenius and spectral norms of a matrix \mathbf{x} are denoted by $\|\mathbf{x}\|_F$ and $\|\mathbf{x}\|_2$ respectively. $E\{\cdot\}$ denotes the expectation of random variables within the brackets. A Gaussian stochastic variable o is then denoted by $o \sim \mathcal{CN}(r, q)$, where r is the mean and q is the variance. Furthermore, a random vector \mathbf{x} having the proper complex Gaussian distribution of mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ is indicated by $\mathbf{x} \sim \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\det(\pi\boldsymbol{\Sigma})} e^{-(\mathbf{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$, for simplicity we refer to $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ as $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

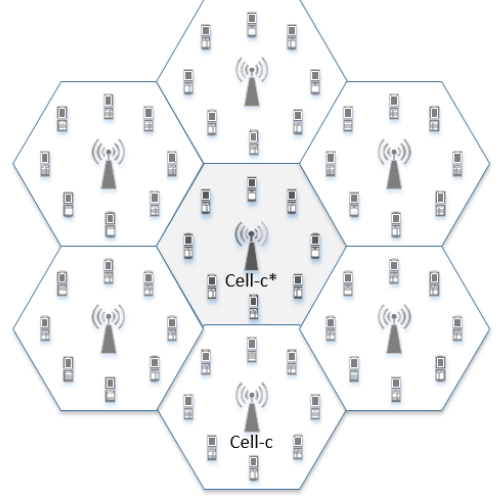


Fig. 1: Illustration of the system model of a multi-cell multi-user massive MIMO.

II. MASSIVE MIMO SYSTEM MODEL

Consider a time division duplexing (TDD) multi-cell massive MIMO system with C cells as shown in Fig. 1. Each cell comprises M antennas at the BS and N single-antenna mobile stations. To improve the spectral efficiency, orthogonal frequency division multiplexing (OFDM) is adopted [11],[12].

At the beginning of the transmission, all mobile stations in all cells synchronously transmit V OFDM pilot symbols to their serving base stations. The v -th pilot symbols of user n in the c -th cell is given by $\mathbf{X}_c^n[v] = [X_c^n[v, 1] X_c^n[v, 2] \dots X_c^n[v, K]]^T$, where K is the number of subcarriers. Let $H_{c^*,c,i}^n[v, k]$ denotes the uplink channel frequency response of the n -th user in the c -th cell sent by the i -th antenna of cell c^* at the k -th subcarrier of the v -th OFDM symbol. The received signal $Y_{c^*,i}$ at the i -th antenna element in the cell c^* can be expressed as

$$Y_{c^*,i}[v, k] = \sum_{n=1}^N H_{c^*,c,i}^n[v, k] X_c^n[v, k] + \sum_{c=1, c \neq c^*}^C \sum_{n=1}^N H_{c^*,c,i}^n[v, k] X_c^n[v, k] + W_{c^*,i}[v, k], \quad (1)$$

for $1 \leq i \leq M$ and $1 \leq c \leq C$, where $W_{c^*,i}[v, k]$ is the uplink channel's additive white Gaussian noise (AWGN). The set of equations constituted by (1) for $1 \leq i \leq M$ can be written as

$$\mathbf{Y}_{c^*}[v, k] = \mathbf{X}_{c^*}[v, k] \mathbf{H}_{c^*,c^*}[v, k] + \sum_{c=1, c \neq c^*}^C \mathbf{X}_c[v, k] \mathbf{H}_{c^*,c}[v, k] + \mathbf{W}_{c^*}[v, k], \quad (2)$$

where $\mathbf{Y}_{c^*}[v, k] \in \mathbb{C}^{1 \times M}$ and $\mathbf{W}_{c^*}[v, k] \in \mathbb{C}^{1 \times M}$ are the two row vectors hosting $Y_{c^*,i}[v, k]$ and $W_{c^*,i}[v, k]$ for $1 \leq i \leq M$, respectively, $\mathbf{X}_{c^*}[v, k] \in \mathbb{C}^{1 \times N}$ and $\mathbf{X}_c[v, k] \in \mathbb{C}^{1 \times N}$ are the two row vectors hosting $X_{c^*}^n[v, k]$ and $X_c^n[v, k]$ for $1 \leq n \leq N$,

respectively, while $\mathbf{H}_{c^*,c^*}[v,k] \in C^{N \times M}$ and $\mathbf{H}_{c^*,c}[v,k] \in C^{N \times M}$ are the two matrices having their n -th row and i -th column elements given by $H_{c^*,c^*,i}^n[v,k]$ and $H_{c^*,c,i}^n[v,k]$, respectively. Assuming that the channel is time-invariant during the channel estimation period, we can drop the OFDM symbol index v from $H_{c^*,c,i}^n[v,k]$. Specifically, $\mathbf{H}_{c^*,c,i}^n[k] = \mathbf{H}_{c^*,c,i}^n$ for the channel estimation period, where row vector $\mathbf{H}_{c^*,c,i}^n = [H_{c^*,c,i}^n[1] H_{c^*,c,i}^n[2] \dots H_{c^*,c,i}^n[K]] \in C^{1 \times K}$ represents the link between the n -th user in the c -th cell over all K OFDM subcarriers. OFDM partitions the multipath channel into a set of parallel and independent sub-channels. The fading coefficient of these sub-channel are the discrete-time Fourier transform of the multipath channel taps. Hence, the term $\mathbf{H}_{c^*,c,i}^n$ can be given as [7],[8]

$$\mathbf{H}_{c^*,c,i}^n = \mathbf{h}_{c^*,c,i}^n \mathbf{F}^T, \quad (3)$$

where the row vector $\mathbf{h}_{c^*,c,i}^n = [h_{c^*,c,i}^n(1) h_{c^*,c,i}^n(2) \dots h_{c^*,c,i}^n(L)] \in C^{1 \times L}$ is the CIR between n -th user at c -th cell and the i -th antenna of the serving BS at the c^* th cell, L is the number of the paths, $\mathbf{F} \in C^{N \times L}$ represents the matrix comprising the discrete Fourier transform (DFT) matrix, the elements of which are given by $[e^{-j2\pi*(k-1)(l-1)/K}]$ for $1 \leq k \leq K$ and $1 \leq l \leq L$. Via exploiting the advantage of cyclic prefix we can omit the subcarrier index k to simplify our notation. Assuming the total V consecutive OFDM symbols are dedicated to pilot subcarriers, so the received signal associated with K OFDM symbols over $1 \leq v \leq V$ can be written as

$$\mathbf{Y}_{c^*} = \mathbf{X}_{c^*} \mathbf{H}_{c^*,c^*} + \sum_{c=1, c \neq c^*}^C \mathbf{X}_c \mathbf{H}_{c^*,c} + \mathbf{W}_{c^*}, \quad (4)$$

The channel coefficient is modelled as $h_{c^*,c,i}^n[\ell] = \sqrt{\phi_{c^*,c,i}^n} g_{c^*,c,i}^n[\ell]$ for $1 \leq \ell \leq L$, where $\phi_{c^*,c,i}^n$ captures the path-loss and shadowing (large-scale fading), while the term $g_{c^*,c,i}^n$ is assumed to be independent identical distribution (i.i.d) of unknown random variables with $CN(0, 1)$ (small-scale fading) [3].

The received signal of (4) can be re-written as

$$\mathbf{Y}_{c^*} = \mathbf{X}_{c^*} \mathbf{F} \mathbf{h}_{c^*,c^*}^n + \mathbf{Z}_{c^*}, \quad (5)$$

where the term $\mathbf{Z}_{c^*} = \sum_{c=1, c \neq c^*}^C \mathbf{X}_c \mathbf{F} \mathbf{h}_{c^*,c}^n + \mathbf{W}_{c^*}$ in (5) represents the net sum of inter-cell interference plus the receiver noise. The variance interference σ_I^2 of the inter-cell interference term caused during pilot transmission can be expressed as

$$\sigma_I^2 = E\{(\sum_{c=1, c \neq c^*}^C \mathbf{X}_c \mathbf{F} \mathbf{h}_{c^*,c}^n)(\sum_{c=1, c \neq c^*}^C \mathbf{X}_c \mathbf{F} \mathbf{h}_{c^*,c}^n)^H\}. \quad (6)$$

We define the measurement matrix $\mathbf{A}_{c^*} = \mathbf{X}_{c^*} \mathbf{F}$, then (5) can be rewritten as

$$\mathbf{Y}_{c^*} = \mathbf{A}_{c^*} \mathbf{h}_{c^*,c^*}^n + \mathbf{Z}_{c^*}. \quad (7)$$

If \mathbf{A}_{c^*} has more rows than columns, i.e. the number of the pilot is greater than the number of the path, then (10) is a standard LS problem with the estimated CIR given by

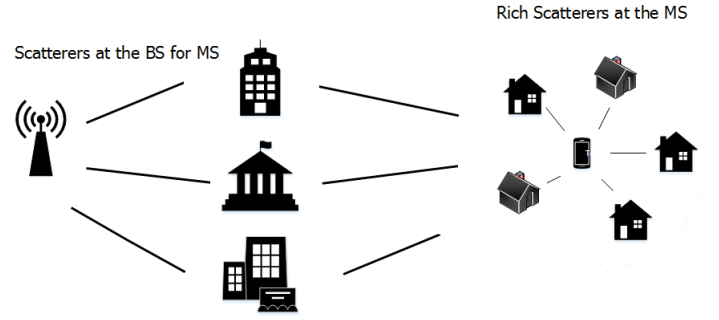


Fig. 2: Illustration of the rich scatterers wireless channel and the resulting channel impulse response is sparse.

$$\hat{\mathbf{h}}_{c^*,c^*} = (\mathbf{A}_{c^*}^H \mathbf{A}_{c^*})^{-1} \mathbf{A}_{c^*}^H \mathbf{Y}_{c^*}. \quad (8)$$

However, we are more interested in the case of the number of the path is greater than the number of the pilot. Based on the physical properties of outdoor electromagnetic propagation, the channel impulse response (CIR) in wireless communications usually possesses several significant channel taps as it shown in Fig. 2, i.e. the CIR are sparse. So, the number of non-zero channel taps is much smaller than the channel length, hence CS techniques can be applied for sparse channel estimation [13]-[15]. This sparsity feature can be exploited to reduce the necessary channel parameters needing to be estimated. In this case, we can address the pilot contamination problem where we can save more pilots and therefore mitigate the pilot contamination and improve the data rate by using fewer pilots than the unknown channel coefficients [13]-[15]. (Based on the compressed sensing theory, the number of measurements should be small compared to the number of signal coefficients, and here the LS and LMMSE are not efficient to estimate the channel.)

It has shown in [], [] and [] that the CIRs of different transmit-receive antenna pairs share a common sparse pattern if $\frac{d}{c} \leq \frac{1}{10B}$, where d is the maximum distance between two BS antennas and c is the speed of light. Therefore, CIRs of different transmit-receive antenna pairs share very similar scatterers, although the corresponding path gains may be quite different

1. Spatial Correlation: Because the scale of the transmit or receive antenna array is very small compared to the long signal transmission distance, channels of different transmit-receive antenna pairs share very similar scatterers. Meanwhile, for most communication systems, the path delay difference from the similar scatterer is far less than the system sampling period. Therefore, CIRs of different transmit-receive antenna pairs share a common sparse pattern, although the corresponding path gains may be quite different [...].

2. Temporal Correlation: For wireless channels, the path delays vary much slowly than the path gains, and the path gains vary continuously [...]. Thus, the channel sparse pattern is nearly unchanged during several adjacent OFDM symbols, and the path gains are also correlated [...]. To take the prac-

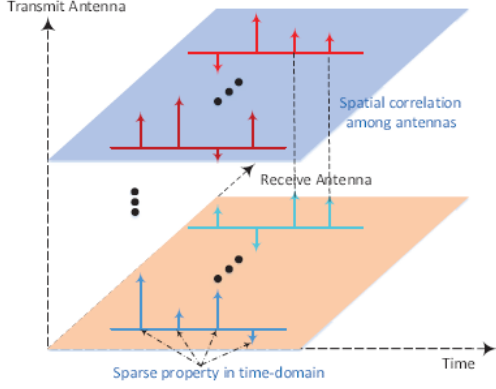


Fig. 3: Illustration of the rich scatterers wireless channel and the resulting channel impulse response is sparse.

tical advantageous of block sparsity, the block structure CS framework is exploited by applying the pattern coupled-SBL approach.

III. SBL-BASED CHANNEL ESTIMATOR

In common literature, channel estimation methods are classified into parametric and Bayesian approaches. A standard parametric approach is the best linear unbiased estimator, which is often referred to as least squares channel estimation. In contrast to parametric methods, the Bayesian approach treats the desired parameters as random variable with a-priori known statistics. Clearly, the a-priori probability density function (PDF) of the channel is assumed to be perfectly known at the receiver [21], [22]. Based on the Bayesian channel estimation philosophy, the estimation of unknown parameters is the expectation of the posterior probabilistic distribution that is proportional to the prior probability and the likelihood of the unknown parameters.

In this section, the pattern-coupled sparse Bayesian learning method is presented in the context of massive MIMO channel estimation. Based on Bayesian channel estimation philosophy, estimated unknown parameters of interest are an expectation of the posterior probability. As such, to obtain the estimated channel, we need to infer the posterior probability of the unknown parameters.

A. Bayesian Inference Model

Following the pattern-coupled sparse Bayesian learning model and based on Bayes' rule [10], the full posterior distribution of \mathbf{h}_{c^*,c^*} over unknown parameters of interest for the problem at hand is proportional to the prior probability and the likelihood of the unknown parameters, that can be computed as

$$P(\mathbf{h}_{c^*,c^*}|\boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{Y}_{c^*}) = P(\mathbf{h}_{c^*,c^*}|\boldsymbol{\alpha})P(\mathbf{Y}_{c^*}|\mathbf{h}_{c^*,c^*}), \quad (9)$$

where $\boldsymbol{\gamma}$ represents the inverse of the net sum of the noise and interference covariance matrices and $\boldsymbol{\alpha}$ are non-negative hyperparameters controlling the sparsity of the channel \mathbf{h}_{c^*,c^*} .

According to probability theory, the term $P(\mathbf{Y}_{c^*}|\mathbf{h}_{c^*,c^*})$ can be written as

$$P(\mathbf{Y}_{c^*}|\mathbf{h}_{c^*,c^*}) = \left(\frac{1}{\sqrt{2\pi}\gamma^{-1}}\right) \exp\left(-\frac{\|\mathbf{Y}_{c^*} - \mathbf{h}_{c^*,c^*}\mathbf{A}_{c^*}\|_2^2}{2\gamma^{-1}}\right), \quad (10)$$

The statistical properties of the sparse multipath structure of the channel is following Gaussian distribution based on the central line theorem [...]. So, the Gaussian prior for each channel coefficient $P(\mathbf{h}_{c^*,c^*}|\boldsymbol{\alpha})$ in the pattern-coupled model is given by

$$P(\mathbf{h}_{c^*,c^*}|\boldsymbol{\alpha}) = \prod_{i=1}^M P(H_{c^*,c^*,i}[v,k]|\alpha_i, \alpha_{i+1}, \alpha_{i-1}) \quad (11)$$

$$= (2\pi)^{-\frac{M}{2}} \prod_{i=1}^M ((\alpha_i, \beta\alpha_{i+1}, \beta\alpha_{i-1}))^{\frac{1}{2}} \exp\left[-\frac{1}{2}(H_{c^*,c^*,i}^n[v,k])^T((\alpha_i, \beta\alpha_{i+1}, \beta\alpha_{i-1}))H_{c^*,c^*,i}^n[v,k]\right],$$

$$i = 1, \dots, M \quad (12)$$

where $0 \leq \beta \leq 1$ is a parameter indicating the pattern relevance between the channel coefficient $H_{c^*,c^*,i}^n[v,k]$ and its neighboring coefficients $\{H_{c^*,c^*,i+1}^n[v,k], H_{c^*,c^*,i-1}^n[v,k]\}$. For $\beta = 0$, the Gaussian prior distribution in (10) is reduced to the prior for the conventional sparse Bayesian learning (which represents the uncorrelated channel scenario).

B. Proposed Algorithm of the Pattern-Coupled Hierarchical Model

We now proceed to perform Bayesian inference for the proposed pattern-coupled SBL-based estimator. The posterior $P(\mathbf{h}_{c^*,c^*}|\boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{Y}_{c^*}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ follows a Gaussian distribution with its mean and covariance given respectively by

$$\boldsymbol{\mu} = \boldsymbol{\gamma}\boldsymbol{\Sigma}\mathbf{A}_{c^*}^H\mathbf{Y}_{c^*}, \quad (13)$$

$$\boldsymbol{\Sigma} = (\mathbf{D} + \boldsymbol{\gamma}(\mathbf{A}_{c^*})^H\mathbf{A}_{c^*})^{-1}, \quad (14)$$

where \mathbf{D} is a diagonal matrix with its i th diagonal element is given by $[\alpha_i, \beta\alpha_{i+1}, \beta\alpha_{i-1}]$, for $i = 1, \dots, M$. The maximum a posterior (MAP) estimate of \mathbf{h}_{c^*,c^*} is the mean of its posterior distribution, i.e.,

$$\hat{\mathbf{h}}_{c^*,c^*} = \boldsymbol{\mu} = ((\mathbf{A}_{c^*})^H\mathbf{A}_{c^*} + \boldsymbol{\gamma}^{-1}\mathbf{D})^{-1}(\mathbf{A}_{c^*})^H\mathbf{Y}_{c^*}. \quad (15)$$

To obtain the term $\hat{\mathbf{h}}_{c^*,c^*}$, we need to jointly estimate the hyperparameters $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, which can be achieved by exploiting the expectation-maximization (EM) approach (we refer interested readers to [10] for detailed derivations). So, the new estimate of $\boldsymbol{\alpha}^{(t+1)}$ and $\boldsymbol{\gamma}^{(t+1)}$ can be given as

$$\alpha_i^{(t+1)} = 10^{-4}/0.5(\hat{\mu}_i^2 + \hat{\Sigma}_{i,i}) + \beta(\hat{\mu}_{i+1}^2 + \hat{\Sigma}_{i+1,i+1}) + \beta(\hat{\mu}_{i-1}^2 + \hat{\Sigma}_{i-1,i-1}) + 10^{-4}, \quad i = 1, \dots, M, \quad (16)$$

where ω_i can be computed as

$$\omega_i = (\hat{\mu}_i^2 + \hat{\Sigma}_{i,i}) + \beta(\hat{\mu}_{i+1}^2 + \hat{\Sigma}_{i+1,i+1}) + \beta(\hat{\mu}_{i-1}^2 + \hat{\Sigma}_{i-1,i-1}), \quad i = 1, \dots, M, \quad (17)$$

$$\begin{aligned} \gamma^{(t+1)} = & M + 2 * 10^{-4} / \|\mathbf{Y}_{c^*} - \mathbf{h}_{c^*,c^*} \mathbf{A}_{c^*}\|_2^2 + (\gamma^{(t)})^{-1} \\ & \sum_i (1 - \hat{\Sigma}_{i,i}(\alpha_i^{(t)} + \beta\alpha_{i-1}^{(t)} + \beta\alpha_{i+1}^{(t)})) \\ & + 2 * 10^{-4}, \quad i = 1, \dots, M, \end{aligned} \quad (18)$$

where φ_i can be given as

$$\varphi_i = 1 - \hat{\Sigma}_{i,i}(\alpha_i + \beta\alpha_{i-1} + \beta\alpha_{i+1}), \quad i = 1, \dots, M. \quad (19)$$

The procedure for implementation of the proposed technique are summarized in algorithms 1.

Algorithm 1 SBL-based Channel Estimator

INPUTS: Pilot Signal \mathbf{X}_{c^*} , observation matrix \mathbf{Y}_{c^*} and the measurement matrix $\mathbf{A}_{c^*} = \mathbf{F}\mathbf{X}_{c^*}$

Initial Configuration:

- 1: Select a specific convergence value ϵ .
- 2: Select a start value for $\alpha^{(t)}$ and $\gamma^{(t)}$
- 3: $t = 0$
- 4: **While** $\|(\hat{\mathbf{h}}_{c^*,c^*})^{(t+1)} - (\hat{\mathbf{h}}_{c^*,c^*})^{(t)}\| \leq \epsilon$ **do**
- 5: Obtain a new estimate for $\alpha^{(t+1)}$ and $\gamma^{(t+1)}$ as in (15) and (17), respectively.
- 6: Compute $\Sigma = (\mathbf{D} + \gamma(\mathbf{A}_{c^*})^H \mathbf{A}_{c^*})^{-1}$
- 7: Compute $\hat{\mathbf{h}}_{c^*,c^*} = \mu = \gamma \Sigma \mathbf{A}_{c^*}^H \mathbf{Y}_{c^*}$
- 8: $t \leftarrow t + 1$
- 8: **end while**

OUTPUTS: Return the Estimated Channel $\hat{\mathbf{h}}_{c^*,c^*}$

C. Optimal Pilot Design

In this section, we will derive the mean square error (MSE) expression for the proposed estimator, as follow

$$MSE = E\|\mathbf{h}_{c^*,c^*} - \hat{\mathbf{h}}_{c^*,c^*}\|_F^2 \quad (20)$$

By applying the results of [—, Chapter 15.8], the mean square error (MSE) expression for the proposed estimator can be given as

$$MSE = \text{tr}\{\mathbf{D} - \mathbf{D}\mathbf{A}_{c^*}^H(\mathbf{A}_{c^*}\mathbf{D}\mathbf{A}_{c^*}^H + \gamma^{-1})^{-1}\mathbf{A}_{c^*}\mathbf{D}\}. \quad (21)$$

Using the the Sherman-Morrison-Woodbury identity,

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{D}\mathbf{A}^{-1}, \quad (22)$$

$$MSE = \text{tr}\{(\mathbf{D} + \gamma\mathbf{A}_{c^*}^H \mathbf{A}_{c^*})^{-1}\} \quad (23)$$

$$= \text{tr}\{(\mathbf{D} + \gamma\mathbf{F}^H(\mathbf{X}_{c^*}^n)^H \mathbf{X}_{c^*}^n \mathbf{F})^{-1}\} \quad (24)$$

we now start designing the optimal pilot signals to improve the channel estimation accuracy by formulating the following optimization problem

$$\begin{aligned} & \underset{\mathbf{X}_{c^*}^n}{\text{minimize}} \quad MSE \\ & \text{subject to} \quad \text{tr}((\mathbf{X}_{c^*}^n)^H \mathbf{X}_{c^*}^n) = P^{UE}. \end{aligned}$$

where $\text{tr}((\mathbf{X}_{c^*}^n)^H \mathbf{X}_{c^*}^n) = P^{UE}$ specifies the power constraint.

The Lagrange multiplier method can be used as follow

$$L(\mathbf{X}_{c^*}^n, \mu) = \text{tr}[(\mathbf{D} + \gamma\mathbf{F}^H(\mathbf{X}_{c^*}^n)^H \mathbf{X}_{c^*}^n \mathbf{F})^{-1}] + \mu(\text{tr}(\sum_{n=1}^N (\mathbf{X}_{c^*}^n)^H \mathbf{X}_{c^*}^n) - P^{UE}) \quad (25)$$

where μ is the Lagrange multiplier.

The optimal training matrix can be obtained by minimizing taking the derivative (22) with respect to $\mathbf{X}_{c^*}^n$ and forcing the result equal to zero.

To that end, using the chain rule we have

$$\begin{aligned} \frac{\partial L(\mathbf{X}_{c^*}^n, \mu)}{\partial (\mathbf{X}_{c^*}^n)^H} &= \frac{\partial \text{tr}[(\mathbf{D} + \gamma\mathbf{F}^H(\mathbf{X}_{c^*}^n)^H \mathbf{X}_{c^*}^n \mathbf{F})^{-1}]}{\partial [(\mathbf{D} + \gamma\mathbf{F}^H(\mathbf{X}_{c^*}^n)^H \mathbf{X}_{c^*}^n \mathbf{F})^{-1}]} \times \\ & \frac{[(\mathbf{D} + \gamma\mathbf{F}^H(\mathbf{X}_{c^*}^n)^H \mathbf{X}_{c^*}^n \mathbf{F})^{-1}]}{\partial \mathbf{X}_{c^*}^n{}^H} + \mu \frac{\partial \text{tr}(\sum_{n=1}^N \mathbf{X}_{c^*}^n{}^H \mathbf{X}_{c^*}^n)}{\partial (\mathbf{X}_{c^*}^n)^H} \\ & - \mu \frac{\partial P^{UE}}{\partial (\mathbf{X}_{c^*}^n)^H} \end{aligned} \quad (26)$$

We are using the following expressions for the matrix derivatives of traces

$$\frac{\partial \text{tr}(\mathbf{Z}(\mathbf{Z})^H)}{\partial (\mathbf{Z})^H} = 2\mathbf{Z}^T \quad (27)$$

$$\begin{aligned} \frac{\partial L(\mathbf{X}_{c^*}^n, \mu)}{\partial (\mathbf{X}_{c^*}^n)^H} &= [-(\mathbf{D} + \gamma\mathbf{F}^H(\mathbf{X}_{c^*}^n)^H \mathbf{X}_{c^*}^n \mathbf{F})^{-2}] \\ & \times [\frac{\partial \text{tr}(\mathbf{D})}{\partial (\mathbf{X}_{c^*}^n)^H} + \gamma \frac{\partial [\mathbf{F}^H(\mathbf{X}_{c^*}^n)^H \mathbf{X}_{c^*}^n \mathbf{F}]}{\partial (\mathbf{X}_{c^*}^n)^H}] + \mu \frac{\partial \text{tr}(\mathbf{F}^H(\mathbf{X}_{c^*}^n)^H \mathbf{X}_{c^*}^n \mathbf{F})}{\partial (\mathbf{X}_{c^*}^n)^H} \end{aligned} \quad (28)$$

$$[-(\mathbf{D} + \gamma\mathbf{F}^H(\mathbf{X}_{c^*}^n)^H \mathbf{X}_{c^*}^n \mathbf{F})^{-2}] \times [0 + \gamma[2\mathbf{X}_{c^*}^n \mathbf{F}]] + \mu 2\mathbf{X}_{c^*}^n \mathbf{F} = 0 \quad (29)$$

$$\gamma\mathbf{F}^H(\mathbf{X}_{c^*}^n)^H \mathbf{X}_{c^*}^n \mathbf{F} = \sqrt{\frac{\gamma}{\mu}} - \mathbf{D} \quad (30)$$

$$(\mathbf{X}_{c^*}^n)^H \mathbf{X}_{c^*}^n = (\mathbf{F}^H)^{-1} \gamma^{-1} \left(\sqrt{\frac{\gamma}{\mu}} - \mathbf{D} \right) \mathbf{F}^{-1} \quad (31)$$

Using the power constrain $\mathbf{X}_{c^*}^n (\mathbf{X}_{c^*}^n)^H = \frac{P^{UE}}{M}$, (23) can be written as

$$L(\mathbf{X}_{c^*}^n, \mu) = \left(\sum_{n=1}^N \phi_{c^*,c^*,i}^n + \frac{P^{UE}}{\Xi^2} \psi \psi^H \right)^{-1} + \mu \left(\frac{P^{UE}}{M} \psi \psi^H \right) \quad (32)$$

$$\frac{1}{\sqrt{\mu}} = \left(\sum_{n=1}^N \phi_{c^*,c^*,i}^n + \Xi^{-2} \frac{P^{UE}}{M} \psi \psi^H \right) \frac{P^{UE}}{M} \psi \psi^H \quad (33)$$

By substituting (25) in (23), we get

$$\sum_{n=1}^N \mathbf{X}_{c^*}^n (\mathbf{X}_{c^*}^n)^H = \left(\sum_{n=1}^N \phi_{c^*,c^*,i}^n + \psi \psi^H \left(\frac{P^{UE}}{M} \right)^2 \Xi \right) - \frac{\Xi^2 \sum_{n=1}^N \phi_{c^*}^n}{\psi \psi^H}$$

Finally, we obtain that the optimal training matrix should satisfy the equation

$$\sum_{n=1}^N \mathbf{X}_{c^*}^n (\mathbf{X}_{c^*}^n)^H = \frac{\Xi}{\sqrt{\mu \psi \psi^H}} - \frac{\Xi^2 \sum_{n=1}^N \phi_{c^*, c^*, i}^n}{\psi \psi^H} \quad (35)$$

D. Enhanced SBL-Based Estimator

In contrast to the proposed SBL-based estimator, the performance of the proposed SBL-based estimator can be improved through the principle of thresholding, which can be applied to retain the most significant taps. The proposed algorithm therefore implements a threshold approach by conserving the channel taps that have energies above a threshold value of ϱ and setting the other taps to zero. The value of ϱ is the energy of the CIR.

E. CRB For SBL-Based Estimator

To quantify the best performance that can be achieved by the proposed algorithm, in this section, we derive the CRB of the pattern-coupled SBL channel estimation. The CRB on the covariance of any estimator $\hat{\theta}$ can be given as $E\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^H\} \geq J^{-1}(\theta)$, where $J(\theta)$ is the Fisher information matrix (FIM) corresponding to the observation f , and can be given as

$$J(\theta) = E\left(\frac{\partial}{\partial \theta} \log l(\theta, f)\right) \left(\frac{\partial}{\partial \theta} \log l(\theta, f)\right)^T, \quad (36)$$

where $l(\theta, f)$ is the likelihood function corresponding to the observation f , parametrized by θ [16].

Theorem 1: The closed form expression of the Bayesian CRB for the proposed SBL can be given as

$$J(\mathbf{h}_{c^*, c^*}) \geq \left(\frac{1}{(\alpha_i, \beta \alpha_{i+1}, \beta \alpha_{i-1})} + \frac{\mathbf{A}_{c^*} (\mathbf{A}_{c^*})^H}{\gamma} \right)^{-1}. \quad (37)$$

$i = 1, \dots, M.$

Proof: See Appendix A.

IV. ACHIEVABLE UPLINK RATE

Given the system model in 2, the BS processes its received signal vector by multiplying it by the conjugate-transpose of the receiver beamforming filter \mathbf{V}_c^H , as follows

$$\mathbf{V}_c^H \mathbf{Y}_{c^*, i} = \mathbf{V}_c^H \sum_{n=1}^N \mathbf{X}_{c^*}^n \mathbf{H}_{c^*, c^*, i}^n + \mathbf{V}_c^H \sum_{c=1, c \neq c^*}^C \sum_{n=1}^N \mathbf{X}_c^n \mathbf{H}_{c^*, c, i}^n + \mathbf{W}_c^H \mathbf{V}_{c^*, i}, \quad (38)$$

The achievable rate can be expressed as

$$R = E\{\log_2(1 + \text{SINR})\}, \quad (39)$$

we can move the expectation to the denominator of the SINR by using the Jensen's inequality that is given as

$$E\{\log_2(1 + \frac{X}{Y})\} \geq \log_2(1 + \frac{E\{X\}}{E\{Y\}}), \quad (40)$$

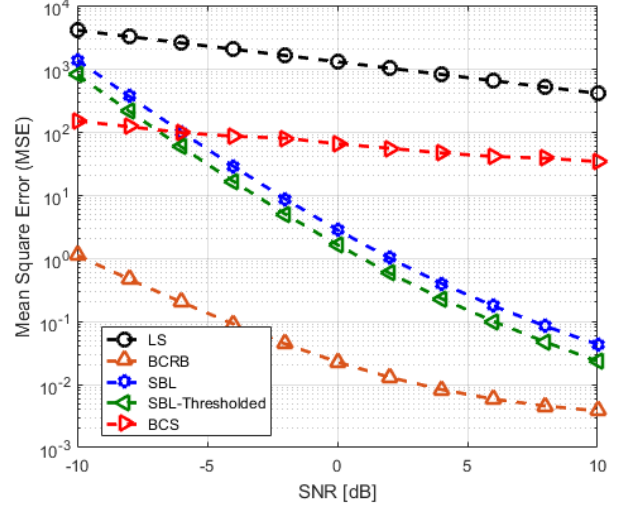


Fig. 4: Relative MSE performance comparison between SBL, Modified SBL, BCS, and the LS versus SNR.

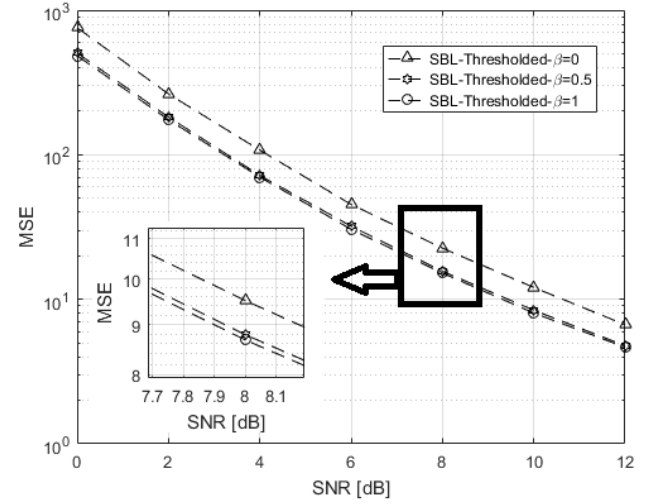


Fig. 5: Relative MSE of the Pattern-Coupled SBL for $\beta = \{0, 0.5 \text{ and } 1\}$.

so, the SINR is the signal to interference and noise ratio and it can be given as

$$\text{SINR} = \frac{|E\{\mathbf{V}_c^H \mathbf{H}_{c^*, c^*, i}^n\}|^2}{\sum_{c=1}^C \sum_{n=1}^N E\{|\mathbf{V}_c^H \mathbf{H}_{c^*, c, i}^n|^2\} - |E\{\mathbf{V}_c^H \mathbf{H}_{c^*, c^*, i}^n\}|^2 + \sigma_v^2 E\{|\mathbf{V}_c|^2\}} \quad (41)$$

The Expectations in the SINR expression (...) are given in closed form by (...) (...), when we employ matched receiver combining (MRC) filter $\mathbf{V}_c = \hat{\mathbf{h}}_{c^*, c^*, i}^n$.

We can compute the numerator as

$$|E\{\mathbf{V}_c^H \mathbf{h}_{c^*, c^*, i}^n\}|^2 = (E\{|\mathbf{V}_c|^2\})^2 \quad (42)$$

$$= \left(\sum_{n=1}^N \phi_{c^*, c^*, i}^n \right)^2 \text{tr}(\psi \sum_{n=1}^N (\mathbf{X}_{c^*}^n)^H \xi^{-1} \sum_{n=1}^N (\mathbf{X}_{c^*}^n) \psi^H) \quad (43)$$

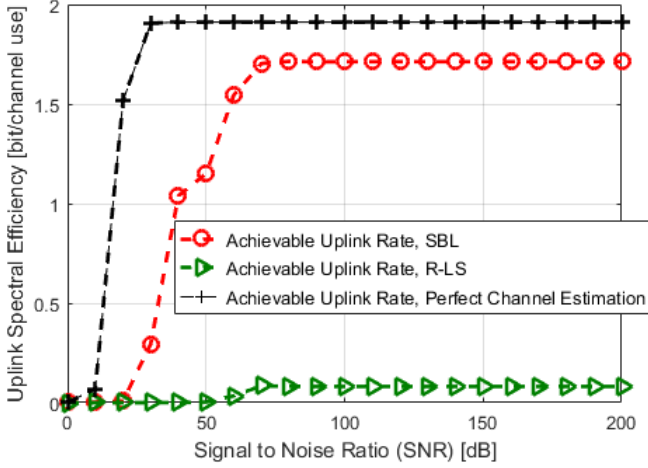


Fig. 6: Relative MSE of the Pattern-Coupled SBL for $\beta = \{0, 0.5 \text{ and } 1\}$.

while the first term on the denominator $E\{|\mathbf{V}_c^H \mathbf{h}_{c^*,c,i}^n|^2\}$ can be computed as

$$E\{|\mathbf{V}_c^H \mathbf{h}_{c^*,c,i}^n|^2\} = \left(\sum_{n=1}^N \phi_{c^*,c,i}^n \right)^2 \text{tr} \left(\boldsymbol{\psi} \sum_{n=1}^N (\mathbf{X}_{c^*}^n)^H \xi^{-1} \sum_{n=1}^N (\mathbf{X}_{c^*}^n) \boldsymbol{\psi}^H \right) \quad (44)$$

while the term $E\{||\mathbf{V}_c^H||^2\}$ can be computed as

$$E\{||\mathbf{V}_c^H||^2\} = \left(\sum_{n=1}^N \phi_{c^*,c,i}^n \right) \text{tr} \left(\boldsymbol{\psi} \sum_{n=1}^N (\mathbf{X}_{c^*}^n)^H \xi^{-1} \sum_{n=1}^N (\mathbf{X}_{c^*}^n) \boldsymbol{\psi}^H \right) \quad (45)$$

By substituting (44), (45) and (46) in (43), the SINR expression becomes

$$\text{SINR} = \quad (46)$$

V. SIMULATION RESULTS

In this section, we conduct experiments to evaluate the performance of the proposed algorithm and compare it to existing methods. The simulation parameters can be summarized as follows, $M = 100$ antennas, $N = 10$ users, $L = 10$ taps and $K = 100$ subcarriers. The simulation scenario is influenced by strong pilot contamination ($\phi_{c^*,c,i}^n = 1$ and $\phi_{c^*,c,i}^n = 0.7$). The simulation results are obtained by averaging over 1000 channel realizations.

Fig. 3 demonstrates the relative MSE performance comparison of least square (LS) with no pilot contamination, SBL-based, thresholded-SBL, Bayesian compressed sensing (BCS) [17] channel estimation techniques along with the BCRB reference line. The results showed that the proposed SBL approach provided significant performance enhancement over LS and BCS with respect to estimation accuracy as a result of exploiting the correlation between the antennas. In addition, the results showed that the thresholding approach strengthened the estimation accuracy of conventional SBL as the CIR possessed so many taps without a significant energy. By setting the threshold and neglecting these taps, a huge portion of the noise and interference from pilot contamination would be eliminated.

Fig. 4 shows the relative MSE performance comparison of SBL-based channel estimation with different settings for $\beta = \{0, 0.5, \text{ and } 1\}$. It can be observed that estimation accuracy is improved when employing the antennas correlation on a large scale.

VI. CONCLUSION

In this paper, we proposed a SBL-based channel estimation algorithm for multi-cell massive MIMO systems. The simulation results revealed that the SBL-based channel estimation algorithm had a tremendous advantage over conventional estimators. Furthermore, the proposed technique can be enhanced by thresholding the CIR to a specific value. In addition, the results demonstrated that the estimation accuracy is enhanced by employing the correlation between antennas on a large scale.

APPENDIX A PROOF OF THEOREM 1

Following (19), we can write the FIM as

$$J(\mathbf{h}_{c^*,c}) \geq E \left(\frac{\partial^2 \log(p_{\mathbf{h}_{c^*,c}/Y_{c^*}}(\mathbf{h}_{c^*,c}, Y_{c^*}))}{\partial^2 \mathbf{h}_{c^*,c}} \right). \quad (47)$$

Based on Bayes' rule in (8), the FIM can be decomposed into two terms

$$\begin{aligned} E \left(\frac{\partial^2 \log(p_{\mathbf{h}_{c^*,c}/Y_{c^*}}(\mathbf{h}_{c^*,c}, Y_{c^*}))}{\partial^2 \mathbf{h}_{c^*,c}} \right) = \\ E \left(\frac{\partial^2 \log(p_{Y_{c^*}/\mathbf{h}_{c^*,c}}(Y_{c^*}, \mathbf{h}_{c^*,c}))}{\partial^2 \mathbf{H}_{c^*,c}} \right) + \\ E \left(\frac{\partial^2 \log(p_{\mathbf{h}_{c^*,c}}(\mathbf{h}_{c^*,c}))}{\partial^2 \mathbf{h}_{c^*,c}} \right). \end{aligned} \quad (48)$$

Which can be expressed in matrix form as

$$\mathbf{J} = \mathbf{J}_D + \mathbf{J}_P. \quad (49)$$

where \mathbf{J} , \mathbf{J}_D and \mathbf{J}_P represent the Bayesian FIM, data information matrix and prior information matrix, respectively.

Using (9), the data information matrix \mathbf{J}_D can be given as

$$\mathbf{J}_D = E \left(\frac{\partial^2 \log(p_{Y_{c^*}/\mathbf{h}_{c^*,c}}(Y_{c^*}, \mathbf{h}_{c^*,c}))}{\partial^2 \mathbf{h}_{c^*,c}} \right) = \frac{\mathbf{A}_{c^*} (\mathbf{A}_{c^*})^H}{\gamma}. \quad (50)$$

Considering (10), the prior information matrix \mathbf{J}_P can be given as

$$\mathbf{J}_P = E \left(\frac{\partial^2 \log(p_{\mathbf{h}_{c^*,c}}(\mathbf{h}_{c^*,c}))}{\partial^2 \mathbf{h}_{c^*,c}} \right) = (\alpha_i, \beta \alpha_{i+1}, \beta \alpha_{i-1})^{-1}. \quad (51)$$

REFERENCES

- [1] Nokia Solutions and Networks, "Technology Vision 2020 White Paper", pp. 116, 2013.
- [2] V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfr ojd, and T. Svensson, "The Role of Small Cells, Coordinated Multipoint, and Massive MIMO in 5G", *IEEE Communications Magazine*, vol. 52, no. 5, pp. 4451, May 2014.
- [3] H. Zhang, S. Gao, D. Li, H. Chen and L. Yang, "On Superimposed Pilot for Channel Estimation in Multi-cell Multiuser MIMO Uplink: Large System Analysis", *IEEE Transaction on Vehicular Technology*, pp. 99, 2015.
- [4] S. Nguyen, "Compressive Sensing for Multi-channel and Large-scale MIMO Networks", PhD. thesis, Dept. of Elect. And Comp. Eng., Concordia Univ., Montreal, Quebec, Canada, 13 Jan 2014.
- [5] C. Qi, G. Yue, L. Wu, Y. Huang and A. Nallanathan, "Pilot Design Schemes for Sparse Channel Estimation in OFDM Systems", *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1493-1505, 2015.
- [6] M. Carlin, P. Rocca, G. Oliveri, F. Viani, and A. Massa, "Directions-of-arrival estimation through Bayesian compressive sensing strategies", *IEEE Trans. Antennas Propagat.*, vol. 61, no. 7, pp. 3828-3838, July 2013.
- [7] R. G. Baraniuk, "Compressive sampling", *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118-124, July 2007.
- [8] Hien Quoc Ngo et al.: "Analysis of the Pilot Contamination Effect in Very Large Multicell Multiuser MIMO Systems for Physical Channel Models", in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech, May 2011.
- [9] J. Fang, Y. Shen, H. Li and P. Wang, "Pattern-Coupled Sparse Bayesian Learning for Recovery of Block-Sparse Signals", *IEEE Transaction on Signal Processing*, vol. 63, no. 2, pp. 360-372, 2015.
- [10] Y. Shen, H. Duan, J. Fang, H. Li, "Pattern-coupled sparse bayesian learning for recovery of block-sparse signals", *Acoustics Speech and Signal Processing (ICASSP) 2014 IEEE International Conference on*, pp. 1896-1900, May 2014.
- [11] J. Zhang, B. Zhang, S. Chen et al., "Pilot contamination elimination for large-scale multiple-antenna aided OFDM systems *IEEE J. Sel. Topics Signal Processing*", pp. 759-772, 2014.
- [12] M. R. Nakhai, "Multicarrier transmission", *IET Signal Process.*, vol. 2, no. 1, pp. 114, Mar. 2008.
- [13] M. Masood, L. Afify, and Tareq Y. Al-Naffouri, "Efficient Coordinated Recovery of Sparse Channels in Massive MIMO," *IEEE Transactions on Signal Processing*, vol. 63, no. 1, Jan. 2015.
- [14] W. Ding, F. Yang, W. Dai and J. Song, "Time-frequency joint sparse channel estimation for MIMO-OFDM systems", *IEEE Commun. Lett.*, vol. 19, no. 1, pp. 58-61, Jan. 2015.
- [15] C. Qi, G. Yue, L. Wu, Y. Huang and A. Nallanathan, "Pilot design schemes for sparse channel estimation in OFDM systems", *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1493-1505, Apr. 2015.
- [16] A. K. Jagannatham and B. D. Rao, "Whitening-Rotation-Based Semi-Blind MIMO Channel Estimation", *IEEE Trans. on Signal Processing*, vol. 54, no. 3, pp. 861-869, 2006.
- [17] H. Al-Salihi and M. R. Nakhai, "Bayesian Compressed Sensing-based Channel Estimation for Massive MIMO Systems", *European Conference on Networks and Communications (EUCNC)*, Athens -EuCNC/EURASIP Student Best Paper Award, 27-30 June 2016.